DOCUMENT RESUME

ED 051 270 TM 000 584

AUTHOR	Nigro, George A.
TITLE	A Strategy to Determine Causal Directions.
PUB DATE	Jan 71
NOTE	9p.; Paper presented in the Session: Statistical Applications, Part II, at the Annual Convention of the National Council on Measurement in Education, New York, New York, February 1971
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Correlation, Educational Research, Hypothesis Testing, *Mathematical Models, *Multiple Regression Analysis, *Research Methodology, *Statistical Analysis
IDENTIFIERS	Beta Index, *Causal Theory

ABSTRACT

A set of mathematical consistencies that forms conditions of inequality in a theorem is summarized, and a strategy for its application with real data is presented. The theorem and strategy are suggested for immediate use by the practitioner seeking cause-effect relationships in a system of variables to cut down guess work and time in analysis and interpretation, especially in light of the masses of data dealt with in the largely nonexperimental or correlational world of education. See also ED 048 362. (Author/AG)

A STRATEGY TO DETERMINE CAUSAL DIRECTIONS

by

GEORGE A. NIGRO

Bristol Community College

Fall River, Massachusetts 02720

January 1971

Presented in the Session: Statistical Applications, Part II
National Council on Measurement in Education
At the Annual Convention
New York City, New York
February 7, 1971

A STRATEGY TO DETERMINE CAUSAL DIRECTIONS

ABSTRACT

This paper summarizes a set of mathematical consistencies that form conditions of inequality in a theorem and presents a strategy for its application with real data, as derived from two studies by the author. The theorem and strategy are suggested for immediate use by the practitioner seeking cause-effect relationships in a system of variables, to cut down guess work and time in analysis and interpretation, especially in light of the requirement to handle masses of data in a largely nonexperimental or correlational world of education.

A STRATEGY TO DETERMINE CAUSAL DIRECTIONS

George A. Nigro

Bristol Community College

A paper presented at the 1969 AERA Annual Meeting[1] exposed a set of inequalities based on the development of a "beta index." This index was proposed as a confirming device for causal directions of three-variable paths in a stepped-regression and correlation analysis from dummy data. These inequalities were completely consistent for causal directions but were inconsistent for acausal directions. The set was generalized to four mathematical statements of asymmetry and proposed as a theorem for a follow-up paper at the 1971 AERA Annual Meeting,[2] which was completed in 1970. The latter study repeated the previous one with one important difference, that horizontal rather than normal distributions were generated and substituted into recursive equations to obtain variables for analysis. The results of both studies were the same. During the course of the latter study and reinforced by a revisit to the previous one, a strategy seemed to emerge for determining causal directions as well as confirming hypothesized directions; that is, regardless of causal directions hypothesized, it appeared possible to generate hypotheses.

Hence, the purpose of this paper is to present such strategy along with a converse of the theorem in revised form.

The educational practitioner often labels his dependent variables as effects in studies, but he often speaks softly in calling his independent variables causes; in fact, he does not dare! If only he had a strategy of determining how his variables operate in a system he chooses, mathematically as well as logically, he would not only find support for his contentions but also cut down guess work and time in analysis and interpretation. Because education is largely a nonexperimental or correlational world and is more and more immersed in masses of data to handle, it would be to the educator's advantage to have such a strategy.

Blalock[3] pulled together materials on causal models, but many years later caution is still extreme to accept his strategy despite increasing suggestions to borrow from fields outside of education. Springboarding from Blalock, the investigator in developing a causal model stumbled into what he named the beta constancy principle: the beta coefficient of an intermediate variable in a causal direction remains relatively constant as other system variables are introduced and controlled in stepped regression, whereas that in the acausal direction changes noticeably. With an assist from Blalock[3] and Driver and Massey[4], the principle served as the seed for the theorem and strategy. The assist, basically, was that the magnitude of the correlations should be smaller between variables furthest removed from one another in causal sequence. These concepts are epitomized in Figure 1 below for a three-variable path and in conditions 1 through 4 of the theorem.

In Figure 1, if variables A, B, and C are related causally, and if other system variables K, as appropriate, are controlled, the rs among the variables would be in ascending order of magnitude $r_{ac \cdot k}$, $r_{ab \cdot k}$, $r_{bc \cdot k}$. Also, β_{cb} would remain relatively constant, though slightly changed, as other variables, variable A mainly, were stepped into regression, whereas β_{ab} would change noticeably.

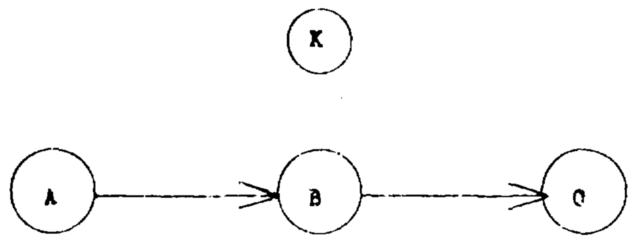

Figure 1. Causal Direction A-B-C

THEOREM FOR CAUSE-EFFECT RELATIONSHIPS

In a three-variable path consisting of variables A, B, and C, with control for other system variables K, if

$$|d\beta_{cb\cdot k/a}| < |d\beta_{ab\cdot k/c}| \qquad (1)$$

where a and c are stepped into regression to obtain beta differentials,

$$|\beta_{ca\cdot kb}| < |\beta_{ac\cdot kb}| \;, \qquad (2)$$

$$|r_{ac\cdot k}| < |r_{ab\cdot k}| < |r_{bc\cdot k}| \;, \qquad (3)$$

$$|r_{ac\cdot kb}| < |r_{ab\cdot kc}| < |r_{bc\cdot ka}| \qquad (4)$$

where k for the first members of the last two conditions are logically derived, then variable A is the initial cause, B the intermediate, and C the final effect.

STRATEGY FOR APPLICATION OF THE THEOREM

1. Assume a set of causal directions in a hypothesized system of variables, then coldly set aside the hypotheses and also disregard intuitive guides until logical analysis is called for.

2. Systematically prepare matrices of the following statistics.

   a. Correlations, both zero-order and partials.
   b. Beta coefficients of all regression equations possible in the system.
   c. Multiple correlations and variances of all possible combinations.

3. Compare blocks of correlations from step 2a by triads, with conditions 3 and 4 as guides, then make tentative decisions as to causal directions. If all appropriate entries yield the same causal directions, hold decision for later analysis with betas. If one of the several triads is inconsistent with directions among other triads, any causal direction should be rejected.

4. Repeat step 3 for each combination of triads in the system.

5. Prepare regression equations to find beta differentials and beta coefficients for end variables, as taken from step 2b, for one selected three-variable path.

   a. Pick a zero-order equation and record the beta coefficient on the tentative intermediate variable, then step in the end variable of the three-variable path selected and record both beta coefficients, a new beta on the intermediate variable and one for the end variable stepped in.

   b. Repeat step 5a for the opposite direction.

   c. Find the algebraic difference between the two betas of the pairs of equations; this step produces two beta differentials for a given three-variable path for opposite directions. Also produced are end-variable betas.

   d. Compare the beta differentials. For condition 1, make a tentative decision as to causal direction.

   e. Compare beta coefficients of end variables stepped in. For condition 2, make a tentative decision as to causal direction.

   f. Repeat steps 5a through 5e with an added variable $\underline{K}$ of the system prior to stepping the end variable of the three-variable path into regression. Any variable $\underline{K}$ is held constant but not considered for either beta-differential or beta-coefficient comparison.

   g. Repeat step 5f until all variables $\underline{X}$ are stepped in, only one at a time.

   h. If all comparisons yield the same causal directions, compare the overall decision as to causal direction to the decision of step 3. If both are the same as to causal direction, record the decision for later logical analysis. If both are inconsistent, any causal direction should be rejected. If any comparision of beta differentials and beta coefficients is inconsistent with directions among others, any causal direction should be rejected. Comparisons are of magnitudes only.

6. Repeat step 5 for each three-variable path, until all three-variable paths have been studied.

7. Prepare a diagram from decisions made, showing connections among variables with arrowhead leaders to show directions of cause-effect. If any link shows mutual causation, arrows in both directions, then, but only then, return to the correlations for a logical analysis; this may be the case when the three variables are confounded.

   a. In the logical analysis, first consideration should be given to the fact that both conditions 3 and 4 are compound inequalities, each consisting of three simple inequalities. Return to the information of step 3 and take any two of the three blocks of correlations; these will already have supplied a definite order of relations for all entries. If there is a correlation entry of the third column which is less than the lesser of the other two columns, inspect the diagram to determine whether it supplies a possible relationship such as the Driver and Massey principle; that is, the relationship between variables furthest removed from one another, in causal sequence, is the least of all relationships.

b. Repeat step 7a until all links are unidirectional, if any link at all.

8. Check patterns of the diagram prepared in step 7 and determine the appropriate regression equations for dependent variables. Be cautious not to include system variables which are not directly related or indirectly related through causal directions.

9. Find total variances accounted for by regression equations of step 8 from step 2c, then find the proportions contributed by each independent variable of the equations. The total from partitions and that from the table of step 2c should be equal with only rounding error. If the total variance accounted for in any equation is low, then possibly one should introduce new variables into the system and repeat the entire process. The caution of step 8, if heeded, should preclude adding false variance just for the sake of variance accounting; it should also preclude variance "caused" in independent variables at the beginning of a causal chain.

10. Compare the diagram of step 7 to that hypothesized and set aside in step 1. Chances are, they are similar. Those directions which were found by the strategy should fall into two categories.

    a. If the directions found are the same as hypothesized, the theorem confirms the hypotheses.

    b. If the directions found were not originally hypothesized, the theorem determines new hypotheses.

APPLICATION OF STRATEGY

The following data have been extracted from the most recent study[2] for steps 3 through 6. Note that Tables 1 and 2 exemplify conditions 3-4 and 1-2, respectively.

Table 1. Correlation Comparisons for Variables 2, 4, and 5, with Controls[a]

Controls[b]	$r_{25.}$	$r_{24.}$	$r_{45.}$	Direction[c]
		For Condition 3[d]		
-	7586	6066	8830	2-5-4
1	8250	7410	8754	2-5-4
3	6339	3805	7919	2-5-4
13	7266	6119	8231	2-5-4
		For Condition 4[e]		
x	5974	2067	8163	2-5-4
x1	5431	0688	6959	2-5-4
x3	5889	2571	7699	2-5-4
x13	4963	0355	6965	2-5-4

a Decimal points precede values listed.
b For partials of rs.
c From comparison of absolute values.
d Control only for variables K.
e Control for third variable in path as well as for variables K; hence, X used.
* At df of 200, r_{01} is greater than or equal to .181[5].

Direction 2-5-4 is the tentative decision for conditions 3 and 4 from the correlation study. Note that r_{25} is less than r_{24}, r_{25} is less than r_{45}, and r_{24} is less than r_{45}, important for later logical analysis.

Regression equations in standard form were next studied for path 254, only those beta coefficients entered as necessary in the table, for variables in the three-variable path 254.

Table 2. Beta Studies for Path 254[a]

	Direction 2-5-4[b]			Direction 4-5-2[c]			Direction[d]	
Other X_i	X_2	X_5	$d\beta_{45}$	$d\beta_{25}$	X_5	X_4	β	$d\beta$
-	-	8830			7586	-		
-	-1489	9959	1129	2533	1.0119	-2869	2-5-4	2-5-4
1	-	7847			8898	-		
1	0489	7411	0436	0760	8138	0969	2-5-4	2-5-4
3	-	7576			6958	-		
3	-1769	8807	1231	2830	9738	-3735	2-5-4	2-5-4
1, 3	-	7610			6886	-		
1, 3	0286	7413	0197	0335	6551	0440	2-5-4	2-5-4

a Decimal points precede values listed except where actually shown.
b Variable 4 is predicted. c Variable 2 is predicted.
d From comparison of absolute values. e Variables in regression equations.

Again direction 2-5-4 is the decision; conditions 1 and 2 are satisfied.

Variables 2, 4, and 5 were purposely selected for presentation here, because direction 2-4-5, not 2-5-4, was the designed causal path in the last study[2]. Obviously, the variables were confounded; but notice how consistent the patterns are. Conditions 3 and 4 were revisited to separate these into three simple inequalities each, as noted above. Some of the entries in the first block of of correlations, r_{25}, are obviously lower in value than some in the second block, r_{24}, all of whose values are less than corresponding values of the third block, r_{45}. But it was also found that comparisons of betas and beta differentials for conditions 1 and 2 consistently produced directions 2-4-5 and 4-2-5, in tables similar to Table 2 above. (No wonder the word "confounded" has an emotional stigma attached to it not unlike Anglosaxonese!)

Returning to the model, repeated in Figure 2 below, then again to the correlation matrix for a logical analysis, it appeared that to satisfy the system best, three inequalities would have to be true; but only one could be true:

For direction 2-5-4, $|r_{24\cdot1}| < |r_{25\cdot4}| < |r_{45\cdot2}|$. (5)

For direction 2-4-5, $|r_{25\cdot4}| < |r_{24\cdot1}| < |r_{45\cdot2}|$. (6)

For direction 4-2-5, $|r_{45\cdot2}| < |r_{24\cdot1}| < |r_{25\cdot4}|$. (7)

Checking the values in Table 1 above, only Inequality 6 is true. This step is admittedly pragmatic, but on approaching logic from the foregoing math base it is considered less so than groping in the dark from personal a priori bias.

The matter of choosing which of the system variables to control brought about much in the way of correlation chess gaming, but nevertheless direction 2-4-5 was the only one of the three considered above that could not be rejected. But the theorem needs revision, to consider conditions 3 and 4 more flexibly by dropping the first member of the two compound inequalities and possibly to add a fifth condition such that

$$|r_{ac\cdot k'}| < |r_{ab\cdot k''}| < |r_{bc\cdot k'''}| \; , \qquad (8)$$

where variables $\underline{K}^{()}$ follow from logical analysis. Such revision remains for further study.

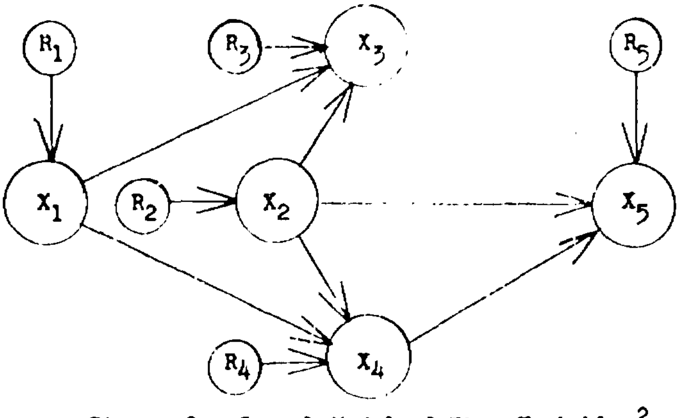

Figure 2. Causal Model of Five Variables[2]
(See Footnote.)

Also found in the previous study by the theorem was direction 1-4-5, as expected, with no confounding whatsoever; no other directions were found causal. A return to the first study revealed the same results about the several paths.

The rest is left as an exercise to students of statistics--literally hours and days.

DISCUSSION AND CONCLUSION

Whether or not hypotheses are made, the strategy to determine causal directions in a system of variables may reveal both relationships and directions. If no directions can be found, then a search for other variables to be brought into the system should be made. Obviously a system of relationships with causal directions seems most significant. Such "system may then serve as a model to conduct an experiment in curriculum, administration, tests, and whatnot or at least further survey research and evaluation of practices. But the theorem must be tried with real data many times and possibly refined to be completely accepted. For the present more computer time must be logged.

REFERENCES

1. Nigro, George A., "A Beta Index to Confirm Causal Directions in a Closed System of Five Variables," a paper presented at AERA Annual Meeting, Feb. 1969; also, ED 031 096, ERIC.

2. _____, "A Theorem to Confirm Causal Directions in a Closed System of Five Variables," a paper presented at AERA Annual Meeting, Feb. 1971; also, ED number pending, ERIC.

3. Blalock, Hubert M., Causal Inferences in Nonexperimental Research, University of North Carolina Press, Chapel Hill, N.C., 1964; p. 73.

4. Driver, H.E., and W.C. Massey, Comparative Studies of North American Indians, The American Philosophical Society, Philadelphia, 1957; pp. 427-434.

5. Arkin, Herbert, and Raymond R. Colton, Tables for Statisticians, 2nd ed., Barnes and Noble, Inc., N.Y., 1965; pp. 24, 155.

FOOTNOTE

The two studies, references 1 and 2, were completed under the Title IV fellowship program, ESEA, PL 89-10 (directed by Dr. John J. Walsh). The model of Figure 2 was used in both studies with the one exception as noted (originally suggested by Dr. Ronald L. Nuttall, also of Boston College).